SUPPLEMENTAL INFO

Structure-system correlation identifies a gene regulatory Mediator submodule

Laurent Larivière[1,5], Martin Seizl[1,5], Sake van Wageningen[2], Susanne Röther[1], Loes van de Pasch[2], Heidi Feldmann[1], Katja Sträßer[1], Steve Hahn[3], Frank C. P. Holstege[2] & Patrick Cramer[1,4]

Supplemental table 1. Crystallographic data

Crystal	SeMet-Med18	Med18	Med8C/18
Data collection			
Space group	C 2 2 2_1	C 2 2 2_1	P 3_2 2 1
Cell parameters			
a (Å)	81.8	81.8	111.0
b (Å)	97.1	97.2	111.0
c (Å)	129.8	130.0	68.6
γ (°)	90	90	120
Wavelength (Å)	0.979	0.990	0.979
Resolution range (Å) [a]	20-2.9	20-2.7	20-2.4
	(3.0-2.9)	(2.80-2.60)	(2.53-2.40)
Completeness (%)	99.6 (97.0)	99.1(94.4)	99.4 (99.4)
Unique reflections	20,243 (1,967)	14,428 (1,344)	19,282 (2,806)
Redundancy	6.9 (4.9)	5.9 (4.7)	11.0 (11.2)
Rsym (%)	8.5 (23.4)	5.6 (31.8)	8.4 (37.9)
<I>/<σ>	18.6 (4.6)	28.3 (3.1)	21.4 (4.8)
Refinement			
Number of residues			224
Number of non-hydrogen atoms			1,885
Number of solvent molecules			62
RMS bond deviation (Å)			0.006
RMS angle deviation (°)			1.293
Ramachandran plot			
(core/allowed/additionally allowed)			89.4/10.1/0.5
R_{cryst}(%)			21.1
R_{free}(%) b			23.1

[a]The numbers in parenthesis correspond to the highest resolution shell

[b]5% of the data were excluded from refinement for free R-factor calculation.

Supplemental table 2. Overrepresented biological processes and associated genes in the yeast deletion strains *med8CΔ*, *med18Δ*, *med20Δ*, *med2Δ* and *med3Δ*. Genes with significantly changed expression levels in all deletion strains of the Med8C/18/20 submodule or Med2 and Med3 are shown in bold.

	Med8C/18/20		**Med2/3**	
GO slim term	**up**	**down**	**up**	**down**
1 - amino acid derivative metabolism	***ARG1***, *ARG3*, *CYS3*, *GLN1*, *HIS4*, *MET2*, *MET28*, ***MET32***, *MHT1*, ***SER3***	*ARG80*, *C**HA1***, ***GCV1***, ***GCV2***, ***GCV3***, ***GLT1,*** *ILV3*, *ILV5*, *LEU9*, ***MET6***, ***MMF1***, *SAM1*	*ADI1*, *ARG1*, *ARG8*, *ARO10*, *ARO9*, ***BAT2***, ***ECM17***, *GDH1*, ***GDH3***, ***MET14,*** *MET16*, ***MET17***, ***MET2***, ***MET28***, ***MET32***, ***MHT1***, ***PUT4***, ***SER3***, *TRP5*, ***UGA1***	***ACO1***, *ALD3*, ***ARG80***, ***CHA1***, ***GAD1***, ***GCV1***, ***GCV2***, ***GCV3***, *GLY1*, ***HIS1***, ***HIS4***, *MET6*, ***MMF1***
3 - carbohydrate metabolic process	***ADR1***, ***AMS1***, *GDB1*, *GLC3*, ***GSC2***,***PGU1***, ***PYC1***	***CAT8,*** ***GAL4***, *GLO4*, *GSY1*, ***HXK1***, ***INO1***, ***KNH1***, ***RKI1***, *SKN1*, ***SOL4***, ***SOR2***, ***TSL1***, *TYE7*	*ADR1*, *AMS1*, ***CIT3***, *DLD1*, *GAL10*, *GAL4*, *GAL7*, ***GSC2***, ***MAL12***, ***MAL32***, *MAL33*, ***NTH2***, *PYC1*	***ACO1***, ***BMH1***, *CTS1*, *EXG1*, *HOR2*, ***HXK1***, *INO1*, ***MDH2***, ***PGM2***, *PSA1*, ***SOL4***, *TPI1*, ***TSL1***
6 - cell wall organization and biogenesis	*FKS3*, *GAS2*, ***GSC2***, *HPF1*, ***SSP2***, *TIP1*	*AMA1*, *CWP1*, *DIT1*, *ECM13*, ***KNH1***, *KTR2*, ***MPC54***, *SKN1*, *SWM1*	***CDA1***, ***CRR1***, ***DIT1***, ***ECM17***, ***ECM8***, *FKS3*, *GIP1*, ***GSC2***, ***OSW2***, ***SMA1***,***SPR28***, ***SPR3***, ***SPS1***, ***SPS22***, ***SSP1***, ***TIP1***	***BMH1***, *EXG1*, *GAS1*, *HPF1*, ***PIR1***, ***PIR3,*** *PSA1*, ***PST1***, ***SMK1***, ***UTR2***, ***ZEO1***
9 - conjugation	***PRM1***, ***PRM5***	*AGA1*, ***ASG7***, ***FAR1***, ***FIG1***, *FUS1*, *FUS3*, *GPA1*, ***PRM2***, *PRM6*, ***PRM7***, *PTP3*, *SST2*, ***STE3***	***PRM2***, *PRM4*, ***PRM5***, ***SPR3***	***CLN2***, ***MFα1***, ***MFα2,*** ***PRM7***
27 - sporulation	*FKS3*, *GAS2*, ***GSC2***, ***SSP2***, *UBX6*	***AMA1***, *DIT1*, ***MPC54,*** ***SPS4***, *SWM1*, *YNL194C*	***CDA1***, ***CRR1***, *FKS3*, *GIP1*, ***GSC2***, ***OSW2***, ***SMA1***, ***SPO20***, ***SPR3***, ***SPS1***, ***SPS19***, ***SPS22***, ***SSP1***	***BMH1***, ***DIT1***, ***SMK1***, *SPS4*
32 - vitamin metabolic process	***PYC1***	*ADH5*, *BIO3*, *BNA1*, ***BNA2***, *BNA4*, ***BNA5***, ***GPD1***, ***RKI1***, *RPI1*, *SNO3*, *SNO4*, ***SNZ3***, ***SOL4***	***BNA1***, ***BNA2***, ***BNA4***, ***BNA5***, *FDH1*, *PYC1*, ***SNO4***, ***THI13***, ***THI4***	***BIO3***, ***BIO5***, ***GPD1***, ***RPI1***, ***SOL4***

Supplemental material and methods

Purification of TAP-Med17

*GAL1::*TAP-*MED17* was generated by integration of *TRP1-GAL1*::TAP at the N-terminus of *MED17* in a RS453 yeast strain. TAP-*MED17* was generated by recombination with Cre recombinase leading to deletion of the *TRP1-GAL1* sequence (Puig et al. 2001). TAP-*MED17 MED8* shuffle was constructed by mating TAP-*MED17* to the *MED8* shuffle strain. This strain was then transformed with pRS315-*MED8* or pRS315-$med8_{1\text{-}189}$ (*med8CΔ*) (Lariviere et al. 2006), followed by shuffling of the pRS316-*MED8* plasmid on 5-fluorotic acid. Med17 and associated proteins were purified from TAP-*MED17 MED8* or TAP-*MED17 med8CΔ* culture. TAP-purification was performed as described (Puig et al. 2001), except that after IgG binding, the column was washed with buffer containing 250 mM NaCl. Copurifying proteins were analysed by 4-12% SDS-PAGE, and bands were analysed by mass spectrometry.

Protein expression and purification

The gene for *Sp* Med18 was cloned into vector pET24b (Novagen), introducing a C-terminal hexahistidine (His) tag. For bicistronic expression of *Sp* Med8C with Med18, the DNA sequence corresponding to residues 180-200 of Med8 was additionally inserted into the vector, together with a second ribosomal binding site as described (Lariviere et al. 2006). Transformed *E. coli* BL21 (DE3) RIL cells (Stratagene) were grown in LB medium at 37 °C to an optical density of 0.5 at 600 nm. Expression was induced with 0.5 mM IPTG for 16 h at 18 °C. For selenomethionine labeling of Med18, expression was carried out as described (Budisa et al. 1995; Meinhart et al. 2003). For protein purification, cells were lysed by sonication in buffer A (20 mM Tris pH 8.0, 150 mM NaCl, 10 mM b-mercaptoethanol). After centrifugation, the supernatant was loaded onto a Ni-NTA column (Qiagen) equilibrated with buffer A. The column was washed with 40 ml of buffer A. Bound proteins were eluted with

buffer A containing 300 mM imidazole. Proteins were further purified by anion exchange chromatography (Mono Q). The column was equilibrated with buffer C (20 mM Tris pH 8.0, 100 mM NaCl, 5 mM dithiothreitol), and proteins were eluted with a linear gradient of 10 column volumes from 100 mM to 1 M NaCl in buffer C. After concentration, the sample was applied to a Superose-12 size exclusion column equilibrated with buffer C containing 150 mM NaCl. Proteins were concentrated to 40 mg ml^{-1} for crystallisation. For *in vitro* transcription assays, *Sc* Med18/20 was expressed using a bicistronic vector constructed as described (Lariviere et al. 2006), except that the His tag was removed. Expression and purification were performed as for *Sp* Med18 and Med8C/18, but Ni-NTA affinity purification was replaced by a precipitation with 30% saturated ammonium sulfate. Med8C/18/20 carries a Strep-tag II and was expressed and purified as described (Lariviere et al. 2006).

Crystallization and X-ray structure determination

Crystals of *Sp* Med18 and of the Med8C/18 heterodimer were grown at 20 °C in hanging drops over reservoirs containing 100 mM Tris pH 8.5, 2 M sodium acetate, and 2 M sodium formate. Crystals were harvested by adding glycerol to a final concentration of 20% (v/v), and were flash-cooled in liquid nitrogen. Diffraction data for native Med18 crystals were collected at 100 K on an ADSC Q210 detector on the ID29 beamline at the ESRF, Grenoble, France. Data for labelled Med18 and the heterodimeric Med8C/18 were collected on a MAR225 detector at the SWISS light Source, Villigen, Switzerland. For labelled Med18, diffraction data were processed with Denzo and Scalepack (Otwinowski and Minor 1996). Program SnB (Weeks and Miller 1999) identified eight selenium sites, which were used for phasing with SOLVE (Terwilliger 2004). Solvent flattening, non-crystallographic symmetry averaging, and building of an initial protein model were done with the program RESOLVE (Terwilliger 2004). The resulting electron density map allowed for building most of Med18 using

TURBO-FRODO (Roussel and Cambillau 1989). The model was refined using the conjugate gradient in the program CNS (Brunger et al. 1998). This model was used for further refinement with the native dataset. Refinement converged at a relatively high free R-factor (32.7%) for unknown reasons, so the free Med18 structure was not deposited. For the dimeric complex Med8C/18, diffraction data were processed with MOSFLM (Leslie 1992) and SCALA (CCP4 1994). The structure was phased by molecular replacement with PHASER (McCoy et al. 2005), which positioned one copy of the partially refined Med18 structure in the asymmetric unit. A model-phased difference Fourier map revealed density for the Med8 C-terminal region, and minor changes in the Med18 structure. After rebuilding, the new model was refined with CNS. The refined structure has a free R-factor of 23.1% and good stereochemistry (Supplemental table 1). The Ramachandran plot showed 89.4% of the residues in the most favoured region and none in disallowed regions.

In vitro transcription

med18Δ and *med8CΔ* strains were generated as described (Lariviere et al. 2006). Nuclear extracts were prepared from three liters of culture as described (www.fhcrc.org/labs/hahn). Plasmid transcription was performed essentially as described (Ranish and Hahn 1991). Transcription reactions were carried out in a 25 μl volume. The reaction mixture contained 100 μg nuclear extract, 150 ng of pSH515 plasmid, 24 ng of Gal4-VP16 activator, 1x transcription buffer (10 mM HEPES pH 7.6, 50 mM potassium acetate, 0.5 mM EDTA, 2.5 mM magnesium acetate), 2.5 mM DTT, 192 μg of phosphocreatine, 0.2 μg of creatine phosphokinase, 10 U of RNase inhibitor (GE Healthcare) and 100 μM nucleoside triphosphates (NTPs). The reaction was incubated at room temperature for 40 min and then stopped with 180 μl of 100 mM sodium acetate, 10 mM EDTA, 0.5% sodium dodecyl sulfate, 17 μg of tRNA/ml. Samples were extracted with phenol-chloroform and precipitated with ethanol. Transcripts were analyzed by primer extension essentially as described (Ranish and

Hahn 1991). Instead of the ^{32}P-labeled lacI oligo, 0.125 pmol of a fluorescently labeled 5'-FAM-oligo was used. Quantitation was performed with a Typhoon 9400 and the ImageQuant Software (Amersham Biosciences).

Gene expression profiling and analysis

All strains except *med8CΔ* are as described (van de Peppel et al. 2005). *med8CΔ* is isogenic to S288c. Truncation was made by using the *adh1* terminator from pFA6a-13myc-kanMX6 (Longtine et al. 1998). All experiments were performed in SC medium with 2% glucose. For microarray analysis, two independent colonies were inoculated and overnight cultures were diluted in fresh medium to an optical density of 0.15 at 600 nm (OD_{600}) (60 ml cultures, 250 rpm shaking incubator 30°C). Cells were harvested by centrifugation (4000 rpm, 3 min) at an OD_{600} of 0.6, and pellets were frozen in liquid nitrogen. The RNA reference was obtained as described (van de Peppel et al. 2005). Total RNA was prepared by hot phenol extraction. Additionally, RNA was treated with DNase (Qiagen) and cleaned up using the RNAeasy kit (Qiagen). We amplified mRNA by *in vitro* transcription using T7 RNA polymerase on 1 μg of total RNA. During *in vitro* transcription, 5-(3-aminoallyl)-UTP (Ambion) was incorporated into the single-stranded cRNA. Cy3 or Cy5 fluorophores (Amersham) were coupled to 3 μg of cRNA. Before hybridization, free dyes were removed using RNAClean (Agencourt), and the efficiency of cDNA synthesis and dye incorporation was measured using a spectrophotometer (SpectraMax190, Molecular Devices).

C6-amino-linked oligonucleotides (70 nucleotides in length), the Yeast Genome ArrayReady (Operon) were purchased from Qiagen and were printed on Codelink slides following manufacturers instructions (GE Healthcare) with a MicroGrid II (Apogent Discoveries) using 48-quill pins (Microspot2500; Apogent Discoveries). Each gene is represented twice and the arrays additionally contained 2838 control features for external control normalization and QC (van de Peppel et al. 2003). From each sample, 2 μg cRNA (with a specific activity of 2–6%

dye-labeled nucleosides) was hybridized, together with 2 μg cRNA from the reference, for 16–20 h at 42 °C (Agilent microarray hybridisation chamber). After scanning (G2565AA Agilent scanner, 100% laser power, 30% photomultiplier tube), raw data was extracted with Imagene 7.5 (Biodiscovery). After image quantification and local background subtraction, all negative values were replaced with the standard deviation of the local background. Print-tip with a span of 0.4 was applied to normalize genes (Yang et al. 2002). After averaging of duplicate spots for each gene on the array, data were visualized and clustered with GeneSpring 7.2 (SiliconGenetics). For each mutant individually, the replicate profiles were compared to the replicate wt profiles through the common reference with ANOVA (R/MAANOVA version 0.98-7 http://cran.r-project.org/src/contrib/Descriptions/maanova.html). In a fixed effect analysis, sample, array, spot, and dye effects were modeled. p values were determined by a permutation F2 test in which residuals were shuffled 5000 times, Benjamini-Hochberg multiple-testing correction was applied. Genes with $p < 0.05$ and an average fold change over the four measurements of at least 1.7 were considered significant. Pearson's correlation was calculated in Microsoft Excel. For each pair of deletion strains the respective lists of significantly changed genes were merged and the respective correlation coefficient (R value) was calculated.

Overrepresented biological processes for genes with significant expression changes were determined using GO Slim Mapper (http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl) based on the GO database (Ashburner et al. 2000). GO Slim Mapper distributes genes to 32 different biological processes according to their GO annotation. Additionally it provides the corresponding genomic background level for each process. All genes categorized with the GO term "Biological process" or "not mapped" were listed together as "not annotated".

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**(1): 25-29.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica* **54**(Pt 5): 905-921.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *European journal of biochemistry / FEBS* **230**(2): 788-796.

CCP4. 1994. The CCP4 suite: programs for protein crystallography. *Acta crystallographica* **50**(Pt 5): 760-763.

Lariviere, L., Geiger, S., Hoeppner, S., Rother, S., Strasser, K., and Cramer, P. 2006. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* **13**(10): 895-901.

Leslie, A.G.W. 1992. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**(10): 953-961.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. 2005. Likelihood-enhanced fast fast translation functions. *Acta crystallographica* **61**: 458-464.

Meinhart, A., Blobel, J., and Cramer, P. 2003. An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem* **278**(48): 48267-48274.

Otwinowski, Z. and Minor, W. 1996. Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzym* **276**: 307-326.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**(3): 218-229.

Ranish, J.A. and Hahn, S. 1991. The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J Biol Chem* **266**(29): 19320-19327.

Roussel, A. and Cambillau, C. 1989. Turbo-FRODO. 77-78.

Terwilliger, T. 2004. SOLVE and RESOLVE: automated structure solution, density modification and model building. *Journal of synchrotron radiation* **11**(Pt 1): 49-52.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F.C. 2003. Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO reports* **4**(4): 387-393.

van de Peppel, J., Kettelarij, N., van Bakel, H., Kockelkorn, T.T.J.P., van Leenen, D., and Holstege, F.C.P. 2005. Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol Cell* **19**(4): 511-522.

Weeks, C.M. and Miller, R. 1999. The design and implementation of *SnB* v2.0. *J Appl Cryst* **32**: 120-124.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**(4): e15.